Visualizing Voice Characteristics with Type Design in Closed Captions for Arabic

Tim Schlippe*, Shaimaa Alessai†, Ghanimeh El-Taweel‡, Matthias Wölfel§ and Wajdi Zaghouani†
*IUBH University of Applied Sciences
†Hamad Bin Khalifa University
‡One Eighty Degrees
§Karlsruhe University of Applied Sciences

***Abstract*—Diversification of fonts in video captions based on the voice characteristics, namely loudness, speed and pauses, can affect the viewer receiving the content. This study evaluates a new method, *WaveFont*, which visualizes the voice characteristics for captions in an intuitive way. The study was specifically designed to test captions, which aims to add a new experience for Arabic viewers. The results indicate that our visualization is comprehensible and acceptable and provides significant added value—for hearing-impaired and non-hearing impaired participants: Significantly more participants stated that *WaveFont* improves their watching experience more than standard captions.**

***Keywords*-closed captions; subtitles; accessibility; speech processing; typography; type design; responsive type; Arabic**

I. Introduction

The number of people watching videos with captions and subtitles is growing: On social media people usually watch videos muted. According to Facebook, 85% of the active users (2 billion) watch videos without sound[1]. According to the World Health Organization, 466 million people (5% of the world's population) are considered as hearing-impaired[2]. On Amazon Prime Video 30% of all viewers turn on closed captions, while 80% of those are not hearing-impaired. As stated in Article 19 of the Universal Declaration of Human Rights[3], issued by the United Nation, "everyone has the right [...] to seek, receive and impart information and ideas through any media and regardless of frontiers".

Conventional subtitles have not evolved for decades. They still reflect *what* is spoken—not *how* it is spoken—i.e. no information about loudness, intonation, pauses, lengths and emotions. They miss out on the opportunity to convey the emotions in the utterances to the audience through typeface, font shape, and size of the font. These elements could be important in keeping additional information present in verbal communication. Moreover, this "emotional gap", experienced by viewers, highlights a significant obstacle to current captions and subtitles; especially when used by the hard-of-hearing and deaf [1]. A further potential is to help children with dyslexia better spell words by 'illustrating' the words through the use of different fonts in the captions [2].

[1]https://digiday.com/media/silent-world-facebook-video
[2]https://www.who.int/news-room/fact-sheets
[3]https://www.un.org/en/universal-declaration-human-rights

We visualize voice characteristics with type design in captions for Arabic. The Arabic script is the second-most widely used writing system in the world by the number of countries using it and the third by the number of users, after the Latin and Chinese scripts [3]. The Arabic language is different from Latin languages in the fact that it is written and read from right to left. According to their position the character's form varies in the word: initial, median, final and insulated. Arabic characters are usually in a connective form. In most cases, the characters transcribe consonants or consonants with vowels. Arabic diacritics located below and above each character within a word are vowelization marks. However, the diacritics are usually absent leaving it to readers to infer an appropriate vowel. The print sector had rather traditional fonts similar to *Simplified Arabic*, which are also used for subtitling. In recent years, however, there have been more and more aesthetically different fonts like *DINArabic*, which was created in 2016 and of which there are also more narrow fonts. In our work we compare the use of the two font families for captioning.

Subtitles in the Arab world are customarily in Modern Standard Arabic (MSA) since they are a translation of a foreign language. As for dialects, the primary approach when captioning, would translate the dialect into MSA. However, for the purpose of this study, to retain the original audible text and the structure as it is, the dialect is further retained within the caption and we do not tackle the challenges of mapping the words spoken in Arabic dialect to the words written in MSA.

This study focuses the evaluation of our innovative technology, called *WaveFont*, to visualize the voice characteristics in captions for the Arabic video content. We present a creative way to adapt the shape of every single word based on the particular phonetic features of the spoken reference. This allows additional information to be presented within the written text of the captions. Consequently, the method supports viewers with captions and subtitles in better understanding the video content.

II. Related Work

In this chapter we describe related work in the fields of font development, captioning and subtitling.

A. Font development and representation

To advance over paper and to fully profit from digital requires going beyond typography as a stylistic device which already has a very long tradition and is used in printed books, posters and comics [4]. Kuhn and Hagenhoff [5] state that "digital reading media [has to be treated] as a system of variable and dynamic design elements for text presentation and text accessibility in communicative spheres." Wölfel [6] argues that text has to be rethought as an independent and alterable media and not be directly compared to its paper-based representation as it might not be a good reference. Novel approaches include kinetic typography which is an animation technique to express ideas using text-based video animation [7, 8], responsive type [9] where the shape of each character is adjusted according to properties such as the relative position of the reader, or to improve the acoustic reading experience by triggered synchronized sound events according to the current read word or word sequence using gaze tracking [10].

To improve comprehension and to include non-textual information, such as emotion or prosody in an utterance, into a visual representation, Wölfel et al. [2] proposed *Voice-Driven Type Design* (VDTD) and registered the method as a patent [11]. It adjusts the shape of each single character according to particular acoustic features in the spoken reference. The motivation of a phoneme-to-grapheme adaptation is to better represent the characteristics of *how* it has been spoken besides *what* has been spoken. VDTD maps the three acoustic properties speed, loudness and pitch to the character width, the horizontal and the vertical stroke weights. While VDTD uses custom font technology to allow for the required variations, it cannot easily be integrated within existing typography rendering workflows. To overcome this technical extra effort, Lacerda Pataca and Costa [12] exploit variable font technology, introduced after VDTD has been developed in OpenType's version 1.8 which is now integrated into all major operating systems and browsers. They also change the granularity to syllable level instead of phoneme/grapheme level and use letter slant to indicate prosody. They found that "participants' responses are highly consistent, indicating that it is indeed plausible to use typographic modulations as a way of representing speech expressiveness, or simply prosody". Bessemans et al. [13] investigated how visual coding of prosody (bold if louder, squished—what we refer to as narrow—if faster, etc.) can help children to improve reading prosody. They found that coding verbal information can create an intuitive representation of speech's expressiveness.

B. Captioning and subtitling

In this research we refer to *interlingual* translation as subtitles and *intralingual* translation as captions. Studies in captioning and subtitling include their placement and design [14, 15]. An eye-tracking study indicates that these parameters can affect reading time and the visual perception of the image [16]. Creative subtitling paves the way for overcoming the limitations of traditional subtitling. McClarty [17] cautions the abidance by traditional subtitling norms as they risk "convert[ing] the translator into a mere rule obeying machine." Based on that, El-Taweel [18] investigates the use of creative subtitles through emojis and emoticons. The use of emojis and emoticons furthers the function of standard subtitling, allowing for tone of voice and emotions to be conveyed to the target audience. The study that targeted audiences with a hearing disability in the Arab world shows that employing emojis and emoticons indicating tone of voice, reduces the amount of reading required for each subtitle, and allows more time to register and understand the subtitles, the emotions and tone of voice expressed. However, results show that emojis and emoticons are not universally understood and may convey several meanings.

Traditional captions and subtitles are limited to telling the audience *what* is merely being said instead of *how* it is being said [19]. These methods do not present information beyond verbatim dialogue such as emotional expressions [20] and can lead to communication problems for the receiving audience [13]. Many studies assert the benefits of captions for the viewers, to make the material more understandable to them [21]. Wölfel et al. [2] state that creative captions and subtitles can benefit a wide range of people, not only deaf and hard-of-hearing.

Manual captioning from scratch can be a very time-consuming process. According to the subtitling company 3PlayMedia, it takes on average 5-10 times the duration of the video. Therefore researchers propose systems for computer-assisted captioning [22–25]. Their systems are based on automatic speech recognition and usually contain an automated transcription, a segmentation to determine which transcribed words are displayed in each caption and an auto-synchronization process which sets the timing based on a forced alignment between the audio and the transcription.

To visualize voice characteristics in closed captions *WaveFont* combines methods from automatic speech recognition, subtitling and typography to automatically and intuitively render characteristics from the voice in captions.

III. WaveFont captions

The following sections describe our implementation to visualize information from the voice in captions.

A. Guidelines

When captioning the test clips, we followed Hamad bin Khalifa University's subtitling guidelines for deaf and hard-of-hearing [26]. These guidelines are based on [27] and have similarities to the BBC's guidelines[4]. The choice to not use traditional subtitle guidelines was made since our target participants were both hearing audiences and those with a hearing disability. In [26], tags are added to describe

[4]https://bbc.github.io/subtitle-guidelines

tone of voice or emotions expressed in speech. For example, in a sports clip, the shout at the goal would have been tagged with [shouts]. However, with *WaveFont*'s creative captions, it is displayed in a bold font to indicate loudness.

B. Visualization of voice characteristics

The visualization of the voice characteristics is on the word level, i.e. for each word, the average values of loudness and speed are used to decide which font to use to represent the whole word. Oral interviews have shown that with captions, the word level is preferred to the character level since viewers see each caption only briefly and therefore have only a short time to interpret the presentation. Consequently, it was very important for us to present the characteristics of the voice as intuitively as possible. Similar to [2], we decided to map the voice to the character shape as follows:

- *Loudness*: Producing loudness in speech amplifies the signal and is usually used to attain the attention of a listener. To have the attention of the reader, bolder text is commonly used since it makes it easier and more efficient to scan the text and recognize important keywords [28]. Therefore, we use a thin font for quieter words and a bold font for louder words.
- *Speed*: The processes of information transfer with speech and reading happens within a time period. A reader usually jumps from a part of a word to a next part of a word [29]. Increasing the character width extends this scanning process of the eyes. Thus, we map the speed of the utterance to the character width: We use a narrow font for fast words and a wide font for slow words.

Our mapping is universally understood across cultures, while this is not the case for emojis and emoticons which may convey several meanings [18]. For aesthetic reasons the different fonts are chosen from the same font family. On the one hand, we aim not to have too extreme differences between the fonts, so that the typeface does not look too restless. On the other hand, the fonts need to be different enough to be easily recognized—even by inexperienced viewers on a small screen. Fig. 1 and 2 show an excerpt of Martin Luther King Jr.'s speech with *WaveFont* captions. The combination of the two representations results in four classes. Fig. 3 summarizes the mapping of the acoustic characteristics loudness and speed to its visual representations stroke weight and character width in our four classes. Our system can display different fonts. To analyze two very different font families, we produced *WaveFont* captions for the following two font families: *Simplified Arabic* and *DINArabic*. *Simplified Arabic* represents the traditional fonts, which are usually used for subtitling. *DINArabic* is a more modern font family which was also proposed in [30].

C. Technical implementation

While in [2] a continuous visual representation was used, we decided to use conventional fonts and subtitle formats in this work to represent the information from the voice. This has the following advantages for captioning and subtitling:

- The *WaveFont* captions can be displayed wherever the font is installed on the system.
- The *WaveFont* captions can be displayed with subtitle formats that support font definition.
- While in a continuous visualization usually vector graphics are generated that need to be burned into the video, conventional subtitle files allow the viewer to turn captions on and off.
- Our process can be easily ported to new languages and writing systems without having to create new characters specific to the new language.
- If the *WaveFont* captions are not burned into the video, the video player allows users to customize them.

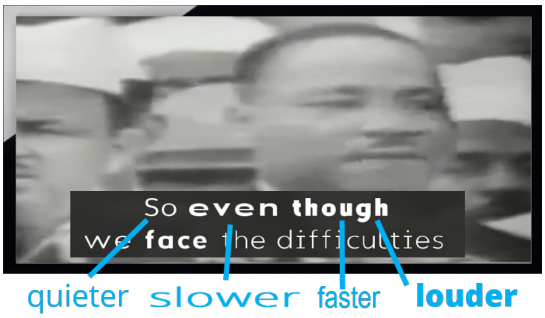

Figure 1. Martin Luther King with English WaveFont captions.

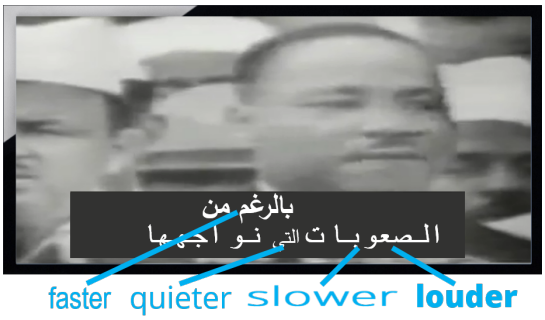

Figure 2. Martin Luther King with Arabic WaveFont captions.

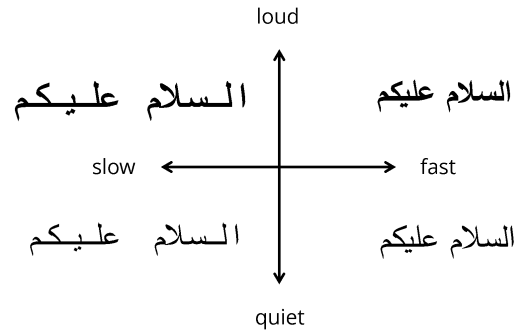

Figure 3. Mapping speech characteristics on text formatting.

To automatically generate *WaveFont*, we use a video and the corresponding caption file as input and apply the following steps:

1) *Extraction of the audio track* from the video.
2) *Segmentation* of the audio track *into smaller audio files* containing spoken utterances based on the time information (start time and end time) of each caption.
3) For each audio speech segment: *Automatic forced alignment* process, which takes the text transcription of the audio speech segment and provides the start and end time of each word in the speech segment.
4) *Acoustic feature extraction*: Provides feature values of loudness and speed of each word. Loudness is based on the signal power. The feature value which represents the speed of each spoken word is computed based on the number of characters and the duration of the uttered word.
5) *Mapping of acoustic features to font classes* based on thresholds.
6) *Type design*: Based on the content of the original subtitle file a new subtitle file is produced that contains font definitions according to the mapped font classes.

A detailed technical description is given in [2] and [11]. For the practical implementation of the last step 6) *Type Design* we had to find solutions for the following challenges:

- Only a few subtitle formats support the display of different fonts in the captions.
- Only a few video players support subtitle formats that allow the display of different fonts.
- Only a few players support the Arabic characters.

We have realized the *WaveFont* captions with the subtitle format SSA/ASS[5] (SubStation Alpha) and used the VLC Player[6] to play our captioned videos.

IV. Experimental setup

In the following sections we describe the selection of videos, the preparation of the survey and the categories which we covered in our questionnaire.

A. Videos

We made a compilation of 76 seconds of footage from the 3 genres *poetry*, *news*, and *sports* in Arabic. *Poetry* was selected since it has always played an essential role in traditional Arabic culture and on TV talent shows for poets are very popular, e.g. Prince of Poets (*Amir al-shūarāa*) and Million's poet (*Shā'ir al-milyūn*). Poets must demonstrate to have a correct diction and to be able to engage the audience. When reciting the poems, emphasis plays a very important role which is why we believe that *WaveFont* can be optimally used here. Additionally, we decided to test the *WaveFont* visualization with the two very popular video genres *news* and *sports*. Sportscasts are full of emotions and can benefit from having the speech characteristics represented. The same applies for news reports. A screenshot of a soccer game with *WaveFont* captions is shown in Fig. 4. The sports clip uses a mixture of MSA and phrases that exist in other dialects and are not specific to one dialect. The poem and the news clip use MSA. However, the poem uses a higher register of MSA than the news clip. The challenges faced when choosing the clips were finding clips in a dialect that is understood in the Arab world and not exclusive to a certain region.

Figure 4. Sports video with Simplified Arabic-WaveFont captions.

B. Questionnaire design

The study examined two different aspects: The first one is regarding the potential benefits of using *WaveFont* for the viewer and the second aspect is the applicability of using *WaveFont* in Arabic captions. Our questionnaire contains the following five parts:

1) Brief introduction of *WaveFont* and general question about the *WaveFont* visualization
2) Playable video with conventional standard captions and related questions
3) Playable video with *WaveFont* captions visualized with the *Simplified Arabic* font and related questions
4) Playable video with *WaveFont* captions visualized with the *DINArabic* font and related questions
5) Personal questions (about general subtitle usage and demographic information)

The font we used for the standard captions in our videos of part 2) is *Simplified Arabic* as these and similar fonts are often used in Arabic captions and subtitles.

V. Experiments and results

159 people (80 female, 79 male) filled out our questionnaire. The participants of our user study were randomly selected volunteers who participated free of charge. Our participants come from 22 countries: Qatar, Oman, UAE, Saudi Arabia, Kuwait, Bahrain, Yemen, Iraq, Palestine, Syria, Jordan, Lebanon, Tunisia, Morocco, Egypt, Pakistan, Turkey, France, Germany, United Kingdom, United States, and Canada. 153 participants' mother tongue is Arabic. 15 subjects were under 25, 65 subjects were in the age range between 25 and 34 years, 49 subjects between 35 and 44,

[5]https://wiki.videolan.org/SubStation_Alpha
[6]https://www.videolan.org/vlc

25 between 45 and 54 and 5 above 55. 32 participants have a hearing disability: 12 are hard-of-hearing and 20 are deaf. 126 participants accessed the questionnaire with a mobile phone, 7 participants with a tablet, and 26 participants with a laptop or PC. As can be seen in Fig. 5, there are participants who often watch subtitled videos, while others never watch with captions or subtitles.

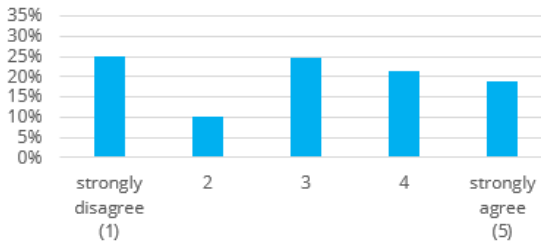

Figure 5. Do you watch subtitled videos? (e.g. on TV or on Social Media)

In order to obtain detailed results, we asked for a score range where it makes sense. The score range follows the rules of a forced choice Likert scale, which ranges from (1) *strongly agree* to (5) *strongly disagree*. In order to also have the comparison to the conventional captions, we asked the participant questions about our two font families for the *WaveFont* captions, but also about the comparison to conventional captions.

A. Intuition and learning effect

We asked the participants at the beginning and again at the end of the questionnaire how intuitive they find *WaveFont*. As shown in Fig. 6, the participants rated this question with an average score of 3.4 at the beginning and at the end with 3.5 after watching our 76-second video. In the beginning, 53% chose *intuitive* or *very intuitive* and in the end it was 56%.

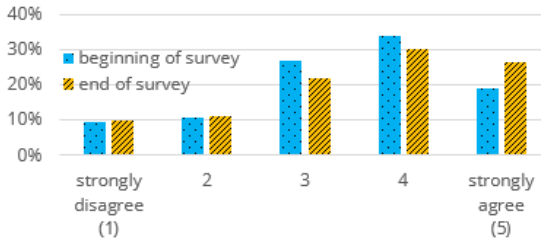

Figure 6. Agree that the WaveFont captions are intuitive to understand.

B. Visualization of voice characteristics

Another goal of our survey was to find out how the visualization of the voice characteristics is received in the captions. Therefore, our participants were asked how comprehensive they found the visualization of loudness and speed after watching our video sequence. Fig. 7 illustrates the evaluation with regard to loudness. While the visualization of loudness in the standard captions was rated with an average score of 2.6, the *WaveFont* visualization of the loudness was rated with 3.4 by the non-hearing impaired (*hearing*) and 3.1 by the hearing-impaired participants (*impaired*), which are relative improvements of 31% and 19%.

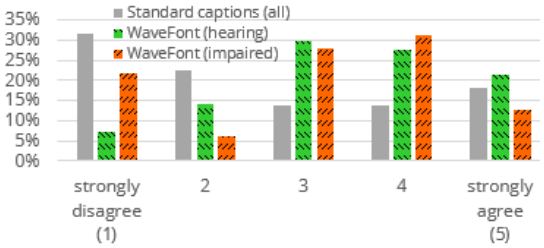

Figure 7. Comprehensible visualization of loudness?

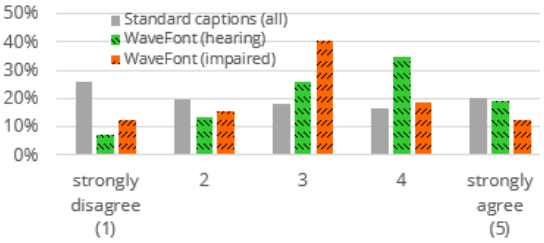

Figure 8. Comprehensible visualization of speed?

Fig. 8 illustrates the evaluation with regard to speed. While the visualization of speed in the standard captions was rated with an average score of 2.6, the *WaveFont* visualization of the speed was rated with 3.4 by *hearing* and 3.0 by *impaired*, which are relative improvements of 31% and 15%.

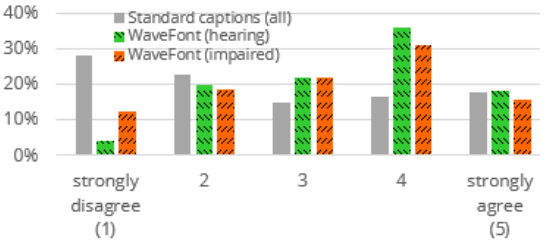

Figure 9. Comprehensible visualization of loudness and speed together?

Fig. 9 demonstrates the combination of loudness and speed. While the visualization in the standard captions was assessed with an average score of 2.8, the *WaveFont* visualization of speed and loudness together was rated with 3.6 by *hearing* and 3.2 by *impaired*, which are relative improvements of 29% and 14%.

The previous results show that the *WaveFont* visualization of the voice characteristics is up to 31% more intuitive than the standard captions. The fact that the results of the hearing-impaired are slightly worse suggests that they cannot distinguish the voice characteristics as accurately–especially those who have never heard.

Since the following results of the participants with and without hearing impairment are comparable without significant differences, we do not list them separately in the upcoming sections but compare the use of the two font families *Simplified Arabic* and *DINArabic* for *WaveFont*.

In our videos we showed *WaveFont* with two classes for loudness and the two classes for speed. We asked the participants if we should add a third class. Fig. 10 reveals that the proportion of participants who prefer a finer distinction in loudness (*soft*, *normal*, *loud*) and those who do not, is comparable, while in speed the larger proportion prefers the two previous classes (*slow*, *fast*).

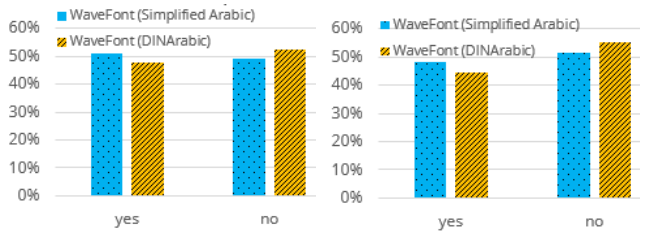

Figure 10. Would you add a third visualization to distinguish between loud, normal and quiet? (*left*) Would you add a third visualization to distinguish between fast, normal and slow? (*right*)

C. Font selection

In the two sections on the two types of *WaveFont* captions, we asked the participants various questions to find out whether our fonts are optimally chosen for displaying the characteristics of the voice or whether changes are desired.

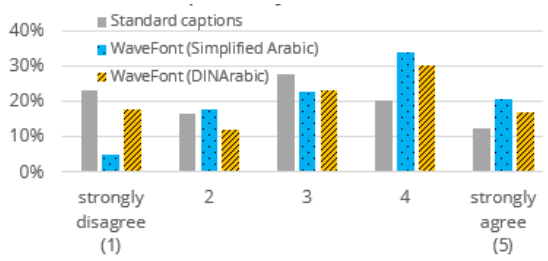

Figure 11. Agree that the fonts are optimally chosen.

As shown in Fig. 11, the question if the fonts are optimally chosen was rated with an average score of 2.8 for the standard captions. For *WaveFont* with *Simplified Arabic* it is 3.5 and for *DINArabic* 3.2, which are relative improvements of 25% and 14%. For *Simplified Arabic*, 55% chose *agree* or *strongly agree* and for *DINArabic* it was 47%.

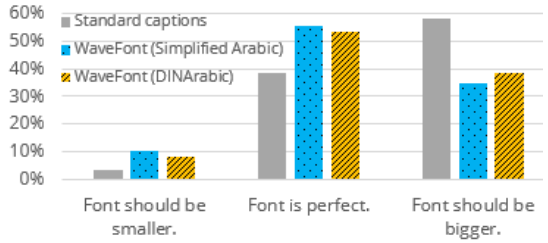

Figure 12. How would you like to change the font size?

Fig. 12 shows that our font size is optimal for most (55% for *Simplified Arabic* and 53% for *DINArabic*), while for standard captions 58% prefer a bigger font.

Fig. 13 and 14 indicate that 52% suggest to display loud words bolder with *Simplified Arabic* and 40% with *DINArabic*. For quiet words, with 53% and 54% the majority is satisfied with the visualization.

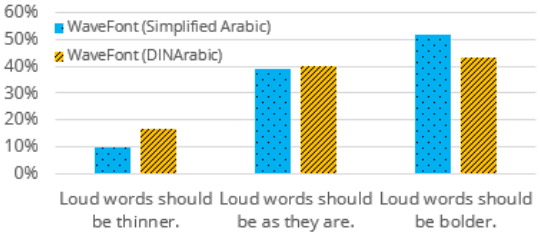

Figure 13. How to display loud words?

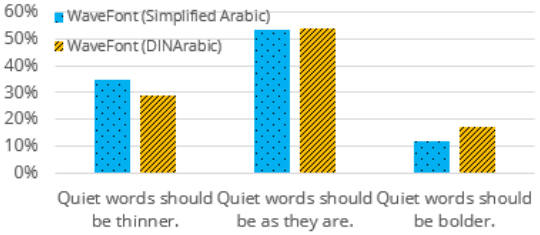

Figure 14. How to display quiet words?

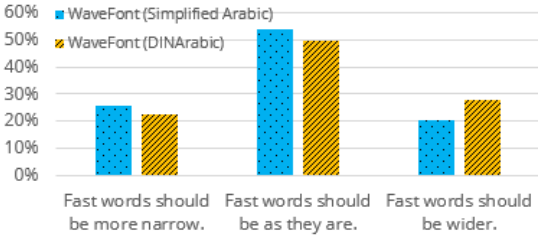

Figure 15. How to display fast words?

As can be seen in Fig. 15 and 16, the majority (in the range of 44% and 54%) is satisfied with the representation of fast and slow words.

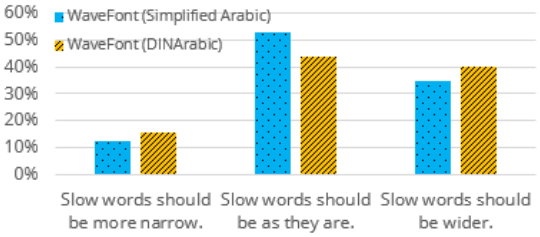

Figure 16. How to display slow words?

When asked which captions improve the TV/cinema experience most, *WaveFont* with *Simplified Arabic* won significantly with 48%, while *DINArabic* was even slightly beaten by standard captions, as shown in Fig. 17. Since *DINArabic* did not perform significantly worse than *Simplified Arabic* in

the questions about comprehensibility, it can be assumed that *Simplified Arabic* scores so much better here as it is closer to the font that viewers are used to. In total, significantly more participants find that *WaveFont* improves their watching experience more than standard captions.

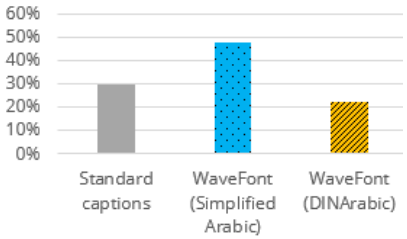

Figure 17. Which captions improve your TV/cinema more?

D. Applications and genres

Fig. 18 shows that 49% in total would prefer a movie where you can turn on *WaveFont* captions. This question was rated with an average score of 3.4. When asked where the participants would like to see *WaveFont*, we get different answers as visualized in Fig. 19. Use cases where more than 30% of the participants agree are: Video-on-demand, TV, social media, live broadcasts and TVs at public places. When it comes to the question in which film genre *WaveFont* should be used, 30% and more of the participants agree on the following genres, as shown in Fig. 20: Comedy, sports, drama, horror, political, historical, thriller, animation, news, explanatory videos, and science fiction.

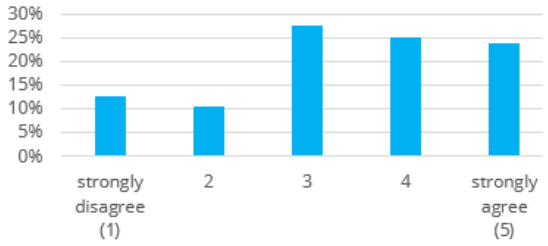

Figure 18. Prefer a movie that optionally displays WaveFont captions?

VI. Conclusion and Future Work

In this paper, we have explored a technology that has the potential to revolutionize the way captions and subtitles are presented, as for the first time, viewers are given information from the voice from which they were previously excluded with traditional captions. Captions are not only interesting for people with a hearing disability. A huge number of people watch videos without sound in social media and closed captions are turned on in noisy environments. Our analysis has shown that the visualization of loudness and speed with *WaveFont* captions is accepted by most people and provides added value–for both non-hearing impaired and hearing-impaired. In our survey with Arabic captions the *WaveFont* visualization with *Simplified Arabic* slightly outperformed the visualization with *DINArabic*. There are

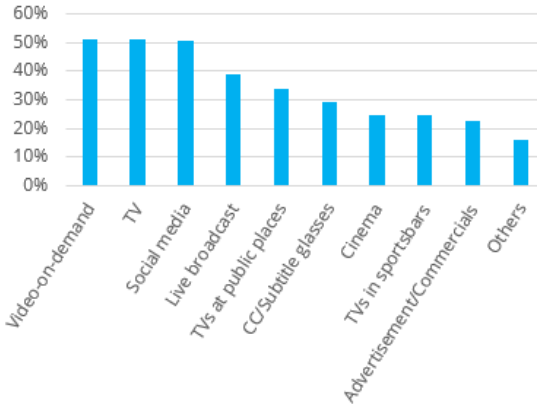

Figure 19. Where do you like to see WaveFont captions?

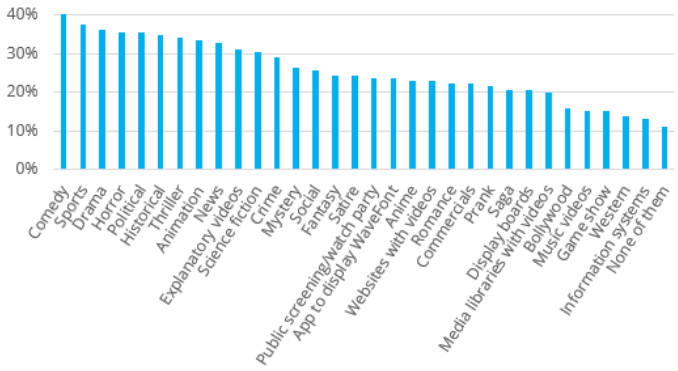

Figure 20. In which genres do you see the potential for WaveFont?

many applications such as video-on-demand, TV, social media, live broadcasts and TVs at public places that can benefit from this technology. Our participants see potential for *WaveFont* in many film genres such as comedy, sports, drama, horror, political, historical, thriller, animation, news, explanatory videos, and science fiction and many more.

Future work may include investigating how to visualize *WaveFont* captions in MSA. This will require a mapping of the speech characteristics of the spoken words in Arabic dialect to the words in the MSA subtitle. To overcome the limited choice of appropriate Arabic font families, the following 3 options could be analyzed to visualize words that are spoken very long as alternatives to using a very wide font: (1) using connective characters with the character *kashida* (-) in between, (2) automatically inserting long vowels, and (3) automatically inserting the diacritics which are usually absent using techniques like [31]. After we have found out that *WaveFont* is also accepted in Arabic, we plan to investigate further languages and writing systems. Further challenges to tackle are to find techniques to transfer the voice characteristics to translated subtitles.

References

[1] J. Ohene-Djan, J. Wright, and K. Combie-Smith, "Emotional Subtitles: A System and Potential Applications for Deaf and Hearing Impaired People," in *CVHI*, 2007.

[2] M. Wölfel, T. Schlippe, and A. Stitz, "Voice Driven Type Design," in *SpeD*, 2015.

[3] "Arabic Alphabet," *Encyclopaedia Britannica*, Retrieved 2020-05-29.

[4] P. Shaw, "Codex: The Journal of Letterforms," *The Menhart Issue. John Boardley*, 2012.

[5] A. Kuhn and S. Hagenhoff, "Kommunikative statt objektzentrierte Gestaltung: Zur Notwendigkeit veränderter Lesekonzepte und Leseforschung für digitale Lesemedien," *Lesen X. 0. Rezeptionsprozesse in der digitalen Gegenwart. Göttingen*, pp. 27–45, 2017.

[6] M. Wölfel, "Rethinking text: Unleashing the full potential of media to provide a better reading experience," *IJACDT*, vol. 7, no. 2, pp. 1–11, 2018.

[7] J. Lee, S. Jun, J. Forlizzi, and S.E. Hudson, "Using Kinetic Typography to Convey Emotion in Text-Based Interpersonal Communication," in *DIS*. 2006, Association for Computing Machinery.

[8] R. Rashid, Q.V. Vy, R.G. Hunt, and D.I. Fels, "Dancing with Words: Using Animated Text for Captioning," *International Journal of Human-Computer Interaction*, vol. 24, pp. 505–519, 2008.

[9] M. Wölfel and A. Stitz, "Responsive Type - Introducing Self-Adjusting Graphic Characters," in *Interspeech*, 2015.

[10] M. Wölfel and D. Hill, "Acoustic Reading Experience: Aligning Sound Events to Text Using Gaze Tracking to Improve Immersion in Reading," in *CERC*, 2017.

[11] T. Schlippe, M. Wölfel, and A. Stitz, "Generation of Text from an Audio Speech Signal," 2018, U.S. Patent 10043519B2.

[12] C. de Lacerda Pataca and P.D.P. Costa, "Speech modulated typography: towards an affective representation model," in *International Conference on Intelligent User Interfaces*, 2020, pp. 139–143.

[13] A. Bessemans, M. Renckens, K Bormans, E. Nuyts, and K. Larson, "Visual prosody supports reading aloud expressively," *Visible Language*, vol. 53, pp. 28–49, 12 2019.

[14] Q.V. Vy and D.I. Fels, "Using Placement and Name for Speaker Identification in Captioning," in *Computers Helping People with Special Needs*, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds., 2010.

[15] A. Brown, R. Jones, M. Crabb, J. Sandford, M. Brooks, M. Armstrong, and C. Jay, "Dynamic Subtitles: The User Experience," in *TVX*, 2015.

[16] W. Fox, "Integrated titles: An improved viewing experience," *Eyetracking and applied linguistics*, 2016.

[17] R. McClarty, "Towards a multidisciplinary approach in creative subtitling," *MonTI: Monografías de Traducción e Interpretación*, pp. 133–153, 2012.

[18] G. El-Taweel, "Conveying Emotions in Arabic SDH: The Case of Pride and Prejudice," 2016, Master Thesis, Hamad Bin Khalifa University.

[19] J. Ohene-Djan, J. Wright, and K. Combie-Smith, "Emotional Subtitles: A System and Potential Applications for Deaf and Hearing Impaired People," in *CVHI*, 2007.

[20] R. Rashid, J. Aitken, and D. Fels, "Expressing Emotions Using Animated Text Captions," 2006, Web Design for Dyslexics: Accessibility of Arabic Content.

[21] M. Gernsbacher, "Video Captions Benefit Everyone," *Policy Insights from the Behavioral and Brain Sciences*, vol. 2, pp. 195–202, 10 2015.

[22] A. Martone, C. Taskiran, and E. Delp, "Automated closed-captioning using text alignment," in *SPIE 5307, Storage and Retrieval Methods and Applications for Multimedia*, 2004.

[23] G. Boulianne, J.-F. Beaumont, M. Boisvert, J. Brousseau, P. Cardinal, C. Chapdelaine, M. Comeau, P. Ouellet, and F. Osterrath, "Computer-assisted closed-captioning of live TV broadcasts in French," in *Interspeech*, 2006.

[24] K. Levin, I. Ponomareva, A. Bulusheva, G. Chernykh, I. Medennikov, N. Merkin, and N. Tomashenko, "Automated closed captioning for Russian live broadcasting," in *Interspeech*, 2014.

[25] S.-E. Tremblay A. Koul, R.G. Kulkarni, "Automated closed captioning using temporal data," 2015, U.S. Patent 9922095B2.

[26] J. Neves, "The (Very) Basics of Enriched Subtitles (for Deaf and Hard of Hearing Audiences) V2," 2019, Hamad bin Khalifa University.

[27] J. Neves, "Audiovisual Translation: Subtitling for the Deaf and Hard of Hearing," 2005, PhD Thesis, Roehampton University.

[28] R. Bringhurst, "The Elements of Typographic Style," *Hartley and Marks Publishers*, vol. 3.2, pp. 55–56, 2008.

[29] G. Unger, "Wie man's liest," *Niggli Verlag*, pp. 63–65, 2006.

[30] S.K. Alessai, "WaveFont for Arabic Video Captioning," 2020, Master Thesis, Hamad bin Khalifa University.

[31] T. Schlippe, T. Nguyen, and S. Vogel, "Diacritization as a Translation Problem and as a Sequence Labeling Problem," in *AMTA*, 2008.